

Interpolation via weighted ℓ_1 minimization

Rachel Ward
University of Texas at Austin

December 12, 2014

Joint work with Holger Rauhut (Aachen University)

Function interpolation

Given a function $f : \mathcal{D} \rightarrow \mathbb{C}$ on a domain \mathcal{D} , interpolate or approximate f from sample values $y_1 = f(u_1), \dots, y_m = f(u_m)$.

Function interpolation

Given a function $f : \mathcal{D} \rightarrow \mathbb{C}$ on a domain \mathcal{D} , interpolate or approximate f from sample values $y_1 = f(u_1), \dots, y_m = f(u_m)$.

Assume the form

$$f(u) = \sum_{j \in \Gamma} x_j \psi_j(u), \quad |\Gamma| = N$$

Approaches:

- ▶ **Standard interpolation:** Choose $m = N$. Find appropriate Γ and sampling points u_1, u_2, \dots, u_m .
- ▶ **Least squares regression:** Choose $N < m$; minimize $\|\mathbf{y} - \mathbf{Ax}\|_2$ where $A_{\ell,j} = \psi_j(u_\ell)$ is the *sampling matrix*.
- ▶ **Compressive sensing methods:** Choose $N > m$. Exploit approximate sparsity of coefficient vector \mathbf{x} to solve the underdetermined system $\mathbf{y} = \mathbf{Ax}$.

Orthonormal systems



L_2 -normalized Legendre polynomials

\mathcal{D} : domain endowed with a probability measure ν .

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, $j \in \Gamma$ (finite or infinite)

$\{\psi_j\}_{j \in \Gamma}$ is an orthonormal system: $\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k}$

Orthonormal systems



L_2 -normalized Legendre polynomials

\mathcal{D} : domain endowed with a probability measure ν .

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, $j \in \Gamma$ (finite or infinite)

$\{\psi_j\}_{j \in \Gamma}$ is an orthonormal system: $\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k}$

Examples:

- ▶ Trigonometric system: $\psi_j(t) = e^{2\pi i j t}$,
 $\mathcal{D} = [0, 1]$, $d\nu(t) = dt$, $\|\psi_j\|_{\infty} \leq 1$.

Orthonormal systems



L_2 -normalized Legendre polynomials

\mathcal{D} : domain endowed with a probability measure ν .

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, $j \in \Gamma$ (finite or infinite)

$\{\psi_j\}_{j \in \Gamma}$ is an orthonormal system: $\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k}$

Examples:

▶ Trigonometric system: $\psi_j(t) = e^{2\pi i j t}$,
 $\mathcal{D} = [0, 1]$, $d\nu(t) = dt$, $\|\psi_j\|_{\infty} \leq 1$.

▶ L_2 -normalized Legendre polynomials L_j :
 $\mathcal{D} = [-1, 1]$, $d\nu(t) = dt$, $\|L_j\|_{\infty} = \sqrt{2j+1}$.

Orthonormal systems



L_2 -normalized Legendre polynomials

\mathcal{D} : domain endowed with a probability measure ν .

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, $j \in \Gamma$ (finite or infinite)

$\{\psi_j\}_{j \in \Gamma}$ is an orthonormal system: $\int_{\mathcal{D}} \psi_j(t) \overline{\psi_k(t)} d\nu(t) = \delta_{j,k}$

► Tensor products of Legendre polynomials:

$\mathcal{D} = [-1, 1]^d$, $\mathbf{j} = (j_1, j_2, \dots, j_d)$, $L_{\mathbf{j}} = \prod_{k=1}^d L_{j_k}$,

$\|L_{\mathbf{j}}\|_{\infty} = \prod_{k=1}^d \sqrt{2j_k + 1}$.

In high-dimensional problems, smoothness is not enough to avoid **curse of dimension** – too local! We will combine smoothness and sparsity.

Smoothness and weights

In general, $\|f\|_{L^\infty} + \|f'\|_{L^\infty}$ promotes smoothness

Consider $\psi_j(t) = e^{2\pi ijt}$, $j \in \mathbb{Z}$, $t \in [0, 1]$

Derivatives satisfy $\|\psi_j'\|_\infty = 2\pi|j|$, $j \in \mathbb{Z}$.

For $f(t) = \sum_j x_j \psi_j(t)$,

$$\begin{aligned}\|f\|_{L^\infty} + \|f'\|_{L^\infty} &= \left\| \sum_j x_j \psi_j \right\|_{L^\infty} + \left\| \sum_j x_j \psi_j' \right\|_\infty \\ &\leq \sum_{j \in \mathbb{Z}} |x_j| (\|\psi_j\|_{L^\infty} + \|\psi_j'\|_{L^\infty}) \\ &= \sum_{j \in \mathbb{Z}} |x_j| (1 + 2\pi|j|) =: \|\mathbf{x}\|_{\omega, 1}.\end{aligned}$$

Smoothness and weights

In general, $\|f\|_{L^\infty} + \|f'\|_{L^\infty}$ promotes smoothness

Consider $\psi_j(t) = e^{2\pi ijt}$, $j \in \mathbb{Z}$, $t \in [0, 1]$

Derivatives satisfy $\|\psi_j'\|_\infty = 2\pi|j|$, $j \in \mathbb{Z}$.

For $f(t) = \sum_j x_j \psi_j(t)$,

$$\begin{aligned}\|f\|_{L^\infty} + \|f'\|_{L^\infty} &= \left\| \sum_j x_j \psi_j \right\|_{L^\infty} + \left\| \sum_j x_j \psi_j' \right\|_\infty \\ &\leq \sum_{j \in \mathbb{Z}} |x_j| (\|\psi_j\|_{L^\infty} + \|\psi_j'\|_{L^\infty}) \\ &= \sum_{j \in \mathbb{Z}} |x_j| (1 + 2\pi|j|) =: \|\mathbf{x}\|_{\omega, 1}.\end{aligned}$$

Weighted ℓ_1 -coefficient norm promotes smoothness. It also promotes sparsity!

Weighted norms and weighted sparsity

Weighted ℓ_1 norm: pay a higher price to include certain indices

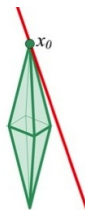
$$\|\mathbf{x}\|_{\omega,1} = \sum_{j \in \Gamma} \omega_j |x_j|$$

New: weighted sparsity:

$$\|\mathbf{x}\|_{\omega,0} := \sum_{j: x_j \neq 0} \omega_j^2$$

\mathbf{x} is called *weighted s -sparse* if

$$\|\mathbf{x}\|_{\omega,0} \leq s.$$



Weighted best s -term approximation error:

$$\sigma_s(\mathbf{x})_{\omega,1} := \inf_{\mathbf{z}: \|\mathbf{z}\|_{\omega,0} \leq s} \|\mathbf{x} - \mathbf{z}\|_{\omega,1}$$

(Weighted) Compressive Sensing

Recover a weighted s -sparse (or weighted-compressible) vector \mathbf{x} from measurements $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m < N$.

Weighted ℓ_1 -minimization

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \mathbf{Az} = \mathbf{y}$$

“Noisy” version

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \|\mathbf{Az} - \mathbf{y}\|_2 \leq \eta$$

(Weighted) Compressive Sensing

Recover a weighted s -sparse (or weighted-compressible) vector \mathbf{x} from measurements $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m < N$.

Weighted ℓ_1 -minimization

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \mathbf{Az} = \mathbf{y}$$

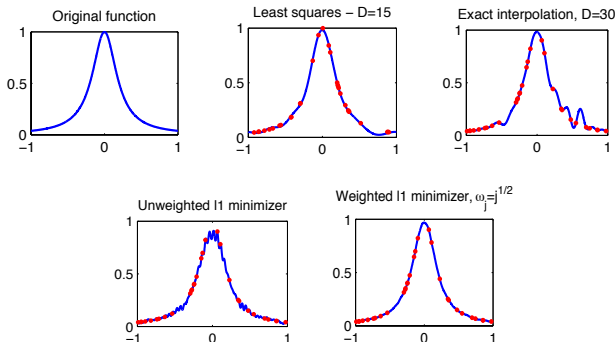
“Noisy” version

$$\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \|\mathbf{Az} - \mathbf{y}\|_2 \leq \eta$$

Keep in mind sampling matrix:

$$A = \begin{pmatrix} \psi_1(u_1) & \psi_2(u_1) & \dots & \dots & \psi_N(u_1) \\ \psi_1(u_2) & \psi_2(u_2) & \dots & \dots & \psi_N(u_2) \\ \vdots & \vdots & & & \vdots \\ \psi_1(u_m) & \psi_2(u_m) & \dots & \dots & \psi_N(u_m) \end{pmatrix}$$

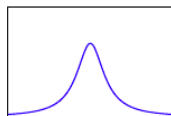
Numerical example - comparing different reconstructions



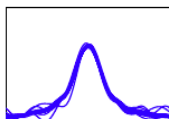
- ▶ $f(u) = \frac{1}{1+25u^2}$. Draw $m = 30$ sampling points u_ℓ i.i.d. from uniform measure on $[-1, 1]$.
- ▶ Interpolate (exactly or approximately) the samples $y_\ell = f(u_\ell)$ by various choices of $\{x_j\}$ in

$$f^\#(u) = \sum_{j=0}^{80} x_j e^{\pi i j u}$$

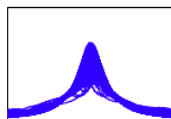
Numerical example - comparing different reconstructions



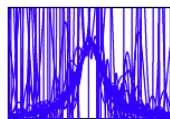
(a) Original function



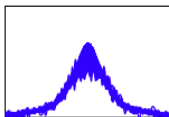
(b) Least squares



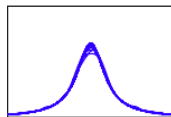
(c) Weighted ℓ_2 , $\omega_j = j^{1/2}$



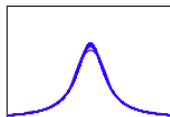
(d) Exact inversion



(e) Unweighted ℓ_1



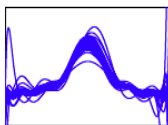
(f) Weighted ℓ_1 , $\omega_j = j^{1/2}$



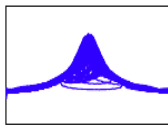
(g) Weighted ℓ_1 , $\omega_j = j$

Different trials correspond to different random draws of $m = 30$ sampling points from uniform measure on $[-1, 1]$.

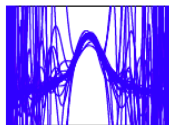
Numerical example - comparing different reconstructions



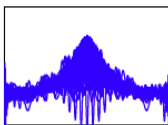
(a) Least squares



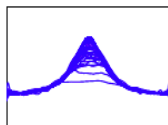
(b) Weighted ℓ_2 , $\omega_j = j^{1/2}$



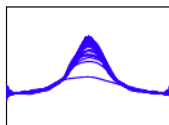
(c) Exact inversion



(d) Unweighted ℓ_1



(e) Weighted ℓ_1 , $\omega_j = j^{1/2}$

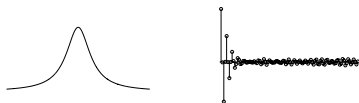


(f) Weighted ℓ_1 , $\omega_j = j$

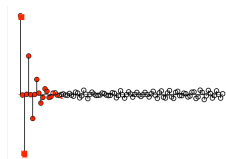
Same experiment, but now interpolating/approximating by Legendre polynomials, and sampling points are from Chebyshev measure on $[-1, 1]$, $d\nu(u) = \frac{du}{\pi\sqrt{1-u^2}}$.

What is going on?

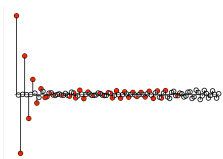
Runge's function and its Legendre polynomial coefficients



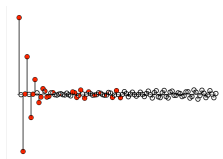
Compare coefficient indices picked up by various reconstruction methods:



Least squares



unweighted l_1 minimization



weighted l_1 minimization

Back to Compressive Sensing setting

Recover a weighted s -sparse (or weighted-compressible) vector \mathbf{x} from measurements $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m < N$.

Weighted ℓ_1 -minimization

$$\min_{\mathbf{z} \in \mathbb{C}^N} \sum_{j=1}^N \omega_j |z_j| \quad \text{subject to } \mathbf{Az} = \mathbf{y}$$

Keep in mind sampling matrix:

$$\mathbf{A} = \begin{pmatrix} \psi_1(u_1) & \psi_2(u_1) & \dots & \dots & \psi_N(u_1) \\ \psi_1(u_2) & \psi_2(u_2) & \dots & \dots & \psi_N(u_2) \\ \vdots & \vdots & & & \vdots \\ \psi_1(u_m) & \psi_2(u_m) & \dots & \dots & \psi_N(u_m) \end{pmatrix}$$

Weighted restricted isometry property (WRIP)

Definition (with H. Rauhut '13)

The weighted restricted isometry constant $\delta_{\omega,s}$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is defined to be the smallest constant such that

$$(1 - \delta_{\omega,s})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_{\omega,s})\|\mathbf{x}\|_2^2$$

for all $\mathbf{x} \in \mathbb{C}^N$ with $\|\mathbf{x}\|_{\omega,0} = \sum_{\ell: x_\ell \neq 0} \omega_\ell^2 \leq s$.

Weighted restricted isometry property (WRIP)

Definition (with H. Rauhut '13)

The weighted restricted isometry constant $\delta_{\omega,s}$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is defined to be the smallest constant such that

$$(1 - \delta_{\omega,s})\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_{\omega,s})\|\mathbf{x}\|_2^2$$

for all $\mathbf{x} \in \mathbb{C}^N$ with $\|\mathbf{x}\|_{\omega,0} = \sum_{\ell: x_\ell \neq 0} \omega_\ell^2 \leq s$.

Unweighted “uniform” restricted isometry property (Candès, Tao '05): $\omega \equiv 1$.

Since we assume $\omega_j \geq 1$, WRIP is weaker than “uniform” RIP

Related: Model-based compressive sensing (Baraniuk, Cevher, Duarte, Hedge, Wakin '10)

Weighted restricted isometry property (WRIP)

Definition (with H. Rauhut '13)

The weighted restricted isometry constant $\delta_{\omega,s}$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is defined to be the smallest constant such that

$$(1 - \delta_{\omega,s}) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_{\omega,s}) \|\mathbf{x}\|_2^2$$

for all $\mathbf{x} \in \mathbb{C}^N$ with $\|\mathbf{x}\|_{\omega,0} = \sum_{\ell: x_\ell \neq 0} \omega_j^2 \leq s$.

Weights allow us to analyze sampling matrices from function systems with unbounded $\|\cdot\|_\infty$ norm. Uniform RIP does not hold for such matrices!

$$\begin{pmatrix} \psi_1(u_1) & \psi_2(u_1) & \dots & \dots & \psi_N(u_1) \\ \psi_1(u_2) & \psi_2(u_2) & \dots & \dots & \psi_N(u_2) \\ \vdots & \vdots & & & \vdots \\ \psi_1(u_m) & \psi_2(u_m) & \dots & \dots & \psi_N(u_m) \end{pmatrix}$$

Weighted RIP of random sampling matrix

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, orthonormal system w.r.t. prob. measure ν . and
 $\|\psi_j\|_\infty \leq \omega_j$.

Sampling points u_1, \dots, u_m taken i.i.d. at random according to ν .
Random sampling matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with entries $A_{\ell j} = \psi_j(u_\ell)$.

Weighted RIP of random sampling matrix

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, orthonormal system w.r.t. prob. measure ν . and
 $\|\psi_j\|_\infty \leq \omega_j$.

Sampling points u_1, \dots, u_m taken i.i.d. at random according to ν .
Random sampling matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with entries $A_{\ell j} = \psi_j(u_\ell)$.
Fix s , and choose N so that $\omega_1, \omega_2, \dots, \omega_N \leq s/2$.

Theorem (with H. Rauhut, '13)

If

$$m \geq C\delta^{-2}s \log^3(s) \log(N)$$

then the weighted restricted isometry constant of $\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies
 $\delta_{\omega,s} \leq \delta$ with high probability.

Weighted RIP of random sampling matrix

$\psi_j : \mathcal{D} \rightarrow \mathbb{C}$, orthonormal system w.r.t. prob. measure ν . and
 $\|\psi_j\|_\infty \leq \omega_j$.

Sampling points u_1, \dots, u_m taken i.i.d. at random according to ν .
Random sampling matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with entries $A_{\ell j} = \psi_j(u_\ell)$.
Fix s , and choose N so that $\omega_1, \omega_2, \dots, \omega_N \leq s/2$.

Theorem (with H. Rauhut, '13)

If

$$m \geq C\delta^{-2}s \log^3(s) \log(N)$$

then the weighted restricted isometry constant of $\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies
 $\delta_{\omega,s} \leq \delta$ with high probability.

Generalizes previous results (Candès, Tao, Rudelson, Vershynin) for uniformly bounded systems, $\|\psi_j\|_\infty \leq K$ for all j .

Recovery via weighted ℓ_1 -minimization

Theorem (with H. Rauhut)

Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ have the WRIP with weights (ω) and $\delta_{\omega, 3s} < 1/3$.
For $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \eta$ let \mathbf{x}^\sharp be a minimizer of

$$\min \|\mathbf{z}\|_{\omega, 1} \quad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta.$$

Then

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_{\omega, 1} \leq C_1 \sigma_s(\mathbf{x})_{\omega, 1} + D_1 \sqrt{s} \eta$$

Generalizes unweighted ℓ_1 minimization results (Candès, Romberg, Tao '06, Donoho '06).

Recovery via weighted ℓ_1 -minimization

Theorem (with H. Rauhut)

Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ have the WRIP with weights (ω) and $\delta_{\omega, 3s} < 1/3$.
For $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \eta$ let \mathbf{x}^\sharp be a minimizer of

$$\min \|\mathbf{z}\|_{\omega, 1} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta.$$

Then

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_{\omega, 1} \leq C_1 \sigma_s(\mathbf{x})_{\omega, 1} + D_1 \sqrt{s} \eta$$

Generalizes unweighted ℓ_1 minimization results (Candès, Romberg, Tao '06, Donoho '06).

Should be of independent interest in structured sparse recovery problems / sparse recovery problems with nonuniform priors on support

Prior work on weighted ℓ_1 minimization

Weighted ℓ_1 norm: $\|\mathbf{x}\|_{\omega,1} = \sum_{j \in \Gamma} \omega_j |x_j|$

Many previous works on the analysis of weighted ℓ_1 minimization!
Focused on the finite-dimensional setting, and with analysis based on unweighted sparsity

- ▶ von Borries, Miosso, and Potes 2007
- ▶ Vaswani and Lu 2009, Jacques 2010, Xu 2010
- ▶ Khajehnejad, Xu, Avestimehr, and Hassibi (2009 & 2010)
- ▶ Friedlander, Mansour, Saab, and Yilmaz 2012, Misra and Parrilo 2013
- ▶ Peng, Hampton, Doostan 2013 - sparse polynomial chaos approximation of high-dimensional stochastic functions

Back to function interpolation

Suppose that $|\Gamma| = N < \infty$.

Given samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ of

$$f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$$

reconstruction amounts to solving $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the sampling matrix $A_{\ell,j} = \psi_j(u_\ell)$

Back to function interpolation

Suppose that $|\Gamma| = N < \infty$.

Given samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ of

$$f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$$

reconstruction amounts to solving $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the sampling matrix $A_{\ell,j} = \psi_j(u_\ell)$

- ▶ Previous analysis suggests: use weighted ℓ_1 -minimization with weights $\omega_j \geq \|\psi_j\|_{L^\infty}$ or $\omega_j \geq \|\psi'_j\|_{L^\infty}$ to recover weighted-sparse or weighted-compressible \mathbf{x} when $m < N$.
- ▶ Choose u_1, \dots, u_m i.i.d. at random according to ν_ψ in order to analyze WRIP of sampling matrix.
- ▶ \mathbf{x} will not be exactly sparse. Measure residual error $f - f^\#$ in which norm? Recall $\|g\|_{L^\infty} + \|g'\|_{L^\infty} \leq \| \|g\|_{\omega,1}$ if $\omega_j \geq \|\psi'_j\|_{L^\infty}$

Interpolation via weighted ℓ_1 minimization

Suppose $f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$, $|\Gamma| = N < \infty$

Set weights $\omega_j > \|\psi_j\|_\infty$

Interpolation via weighted ℓ_1 minimization

Suppose $f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$, $|\Gamma| = N < \infty$

Set weights $\omega_j > \|\psi_j\|_\infty$

Theorem

From $m \asymp s \log^3(s) \log(N)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ where $u_j \sim \nu_\psi$, if \mathbf{x}^\sharp is the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^\Gamma} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}$$

and $f^\sharp(u) = \sum_{j \in \Gamma} x_j^\sharp \psi_j(u)$,

Interpolation via weighted ℓ_1 minimization

Suppose $f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$, $|\Gamma| = N < \infty$

Set weights $\omega_j > \|\psi_j\|_\infty$

Theorem

From $m \asymp s \log^3(s) \log(N)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ where $u_j \sim \nu_\psi$, if \mathbf{x}^\sharp is the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^\Gamma} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \mathbf{Az} = \mathbf{y}$$

and $f^\sharp(u) = \sum_{j \in \Gamma} x_j^\sharp \psi_j(u)$, then with high probability,

$$\|f - f^\sharp\|_\infty \leq \| \|f - f^\sharp\| \|_{\omega,1} \leq C_1 \sigma(f)_{\omega,1}$$

Interpolation via weighted ℓ_1 minimization

Suppose $f(u) = \sum_{j \in \Gamma} x_j \psi_j(u)$, $|\Gamma| = N < \infty$

Set weights $\omega_j > \|\psi_j\|_\infty$

Theorem

From $m \asymp s \log^3(s) \log(N)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ where $u_j \sim \nu_\psi$, if \mathbf{x}^\sharp is the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^\Gamma} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to } \mathbf{Az} = \mathbf{y}$$

and $f^\sharp(u) = \sum_{j \in \Gamma} x_j^\sharp \psi_j(u)$, then with high probability,

$$\|f - f^\sharp\|_\infty \leq \| \|f - f^\sharp\| \|_{\omega,1} \leq C_1 \sigma(f)_{\omega,1}$$

More realistic: $|\Gamma| = \infty$. How to pass to a finite dimensional approximation in a principled way?

Weighted function spaces

$$\mathcal{A}_{\omega,p} = \left\{ f : f(u) = \sum_{j \in \Gamma} x_j \psi_j(u), \quad \|f\|_{\omega,p} := \|\mathbf{x}\|_{\omega,p} < \infty \right\}$$

Interesting range: $0 \leq p \leq 1$

$$\mathcal{A}_{\omega,p} \subset \mathcal{A}_{\omega,q} \text{ if } p < q$$

Solving $m \asymp s \log^3(s) \log(N)$ for s , a Stechkin-type error bound yields

$$\|f - f^\# \|_{L^\infty} \leq \|f - f^\# \|_{\omega,1} \leq C_1 \left(\frac{\log(N)^4}{m} \right)^{1/p-1} \|f\|_{\omega,p}$$

Approximation in infinite-dimensional spaces

Suppose $|\Gamma| = \infty$, $\lim_{|j| \rightarrow \infty} \omega_j = \infty$ and $\omega_j > \|\psi_j\|_\infty$.

Approximation in infinite-dimensional spaces

Suppose $|\Gamma| = \infty$, $\lim_{|j| \rightarrow \infty} \omega_j = \infty$ and $\omega_j > \|\psi_j\|_\infty$.

Theorem

Suppose $f \in \mathcal{A}_{\omega,p}$ for some $0 < p < 1$. Fix s , and set

$$\Gamma_s = \{j \in \Gamma : \omega_j^2 \leq s/2\}.$$

Approximation in infinite-dimensional spaces

Suppose $|\Gamma| = \infty$, $\lim_{|j| \rightarrow \infty} \omega_j = \infty$ and $\omega_j > \|\psi_j\|_\infty$.

Theorem

Suppose $f \in \mathcal{A}_{\omega,p}$ for some $0 < p < 1$. Fix s , and set

$$\Gamma_s = \{j \in \Gamma : \omega_j^2 \leq s/2\}.$$

Draw $m \geq Cs \log^3(s) \log(|\Gamma_s|)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ according to ν_ψ , and let \mathbf{x}^\sharp be the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^{\Gamma_s}} \|\mathbf{z}\|_{\omega,1} \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{m/s} \|f - f_{\Gamma_s}\|_{\omega,1}$$

Put $f^\sharp = \sum_{j \in \Gamma_s} x_j^\sharp \psi_j$.

Approximation in infinite-dimensional spaces

Suppose $|\Gamma| = \infty$, $\lim_{|j| \rightarrow \infty} \omega_j = \infty$ and $\omega_j > \|\psi_j\|_\infty$.

Theorem

Suppose $f \in \mathcal{A}_{\omega,p}$ for some $0 < p < 1$. Fix s , and set

$$\Gamma_s = \{j \in \Gamma : \omega_j^2 \leq s/2\}.$$

Draw $m \geq Cs \log^3(s) \log(|\Gamma_s|)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ according to ν_ψ , and let \mathbf{x}^\sharp be the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^{\Gamma_s}} \|\mathbf{z}\|_{\omega,1} \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{m/s} \|\| f - f_{\Gamma_s} \|\|_{\omega,1}$$

Put $f^\sharp = \sum_{j \in \Gamma_s} x_j^\sharp \psi_j$. Then with high probability,

$$\|f - f^\sharp\|_{L^\infty} \leq \|\| f - f^\sharp \|\|_{\omega,1} \leq C_1 \left(\frac{\log^4(N)}{m} \right)^{1/p-1} \|\| f \|\|_{\omega,p}$$

Approximation in infinite-dimensional spaces

Suppose $|\Gamma| = \infty$, $\lim_{|j| \rightarrow \infty} \omega_j = \infty$ and $\omega_j > \|\psi_j\|_\infty$.

Theorem

Suppose $f \in \mathcal{A}_{\omega,p}$ for some $0 < p < 1$. Fix s , and set

$$\Gamma_s = \{j \in \Gamma : \omega_j^2 \leq s/2\}.$$

Draw $m \geq Cs \log^3(s) \log(|\Gamma_s|)$ samples $y_1 = f(u_1), \dots, y_m = f(u_m)$ according to ν_ψ , and let \mathbf{x}^\sharp be the solution to

$$\min_{\mathbf{z} \in \mathbb{C}^{\Gamma_s}} \|\mathbf{z}\|_{\omega,1} \text{ subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{m/s} \|\| f - f_{\Gamma_s} \|\|_{\omega,1}$$

Put $f^\sharp = \sum_{j \in \Gamma_s} x_j^\sharp \psi_j$. Then with high probability,

$$\|f - f^\sharp\|_{L^\infty} \leq \|\| f - f^\sharp \|\|_{\omega,1} \leq C_1 \left(\frac{\log^4(N)}{m} \right)^{1/p-1} \|\| f \|\|_{\omega,p}$$

**Using greedy alternative such as weighted iterative hard thresholding, do not need to know $\sqrt{m/s} \|\| f - f_{\Gamma_s} \|\|_{\omega,1}$, get same bound

Function approximation in high dimensions

Important: For domains $\mathcal{D} = [-1, 1]^d$, don't want number of measurements m in resulting bound to grow much with d .

Function approximation in high dimensions

Important: For domains $\mathcal{D} = [-1, 1]^d$, don't want number of measurements m in resulting bound to grow much with d .

L^2 -normalized Chebyshev polynomials

$$T_j(u) = \sqrt{2} \cos(j \arccos u), \quad T_0(u) \equiv 1, \quad j \in \mathbb{N},$$

Tensorized Chebyshev polynomials

$$T_{\mathbf{k}}(u) = \prod_{j=1}^d T_{k_j}(u_j) = \prod_{j \in \text{supp } \mathbf{k}} T_{k_j}(u_j), \quad u = (u_j)_{j \geq 1}, \mathbf{k} \in \mathcal{F},$$
$$\mathcal{F} = \{\mathbf{k} = (k_1, k_2, \dots), k_j \in \mathbb{N}\}$$

Function approximation in high dimensions

Important: For domains $\mathcal{D} = [-1, 1]^d$, don't want number of measurements m in resulting bound to grow much with d .

L^2 -normalized Chebyshev polynomials

$$T_j(u) = \sqrt{2} \cos(j \arccos u), \quad T_0(u) \equiv 1, \quad j \in \mathbb{N},$$

Tensorized Chebyshev polynomials

$$T_{\mathbf{k}}(u) = \prod_{j=1}^d T_{k_j}(u_j) = \prod_{j \in \text{supp } \mathbf{k}} T_{k_j}(u_j), \quad u = (u_j)_{j \geq 1}, \mathbf{k} \in \mathcal{F},$$
$$\mathcal{F} = \{\mathbf{k} = (k_1, k_2, \dots), k_j \in \mathbb{N}\}$$

Product probability measure

$$\eta = \bigotimes_{j \geq 1} \frac{du_j}{\pi \sqrt{1 - u_j^2}}, \quad \in \mathcal{D}.$$

$\{T_{\mathbf{k}}\}_{\mathbf{k} \in \mathcal{F}}$ forms orthonormal basis for $L^2(\mathcal{D}, d\eta)$.

Sparse recovery for tensorized Chebyshev polynomials

L^∞ -bound for the tensorized Chebyshev polynomials $T_{\mathbf{k}}$:

$$\|T_{\mathbf{k}}\|_\infty = 2^{\|\mathbf{k}\|_0/2}.$$

Choice of weights:

$$\omega_{\mathbf{k}} = \bigotimes_{j=1}^d (k_j + 1)^{1/2} \geq 2^{\|\mathbf{k}\|_0/2}$$

Encourages both *sparse* and *low-order* tensor products

Sparse recovery for tensorized Chebyshev polynomials

L^∞ -bound for the tensorized Chebyshev polynomials $T_{\mathbf{k}}$:

$$\|T_{\mathbf{k}}\|_\infty = 2^{\|\mathbf{k}\|_0/2}.$$

Choice of weights:

$$\omega_{\mathbf{k}} = \bigotimes_{j=1}^d (k_j + 1)^{1/2} \geq 2^{\|\mathbf{k}\|_0/2}$$

Encourages both *sparse* and *low-order* tensor products

Subset of indices forms a hyperbolic cross:

$$\Lambda_0 = \{\mathbf{k} \in \mathbb{N}_0^d : \omega_{\mathbf{k}}^2 \leq s\}$$

[Kuhn, Sickel, Ulrich, 2014, Cohen, DeVore, Foucart, Rauhut 2011]:

$$|\Lambda_0| \leq e^d s^{2+\log(d)}$$

Apply weighted ℓ_1 theory:

Consider a function

$$f = \sum_{\mathbf{k} \in \Lambda} x_{\mathbf{k}} T_{\mathbf{k}} \quad \text{on } [-1, 1]^d, \quad f \in \mathcal{A}_{\omega, p}$$

- ▶ Fix s , and fix $\Lambda_0 = \{\mathbf{k} \in \mathbb{N}^d : \omega_{\mathbf{k}}^2 \leq s\}$ **Reduced basis**
- ▶ Fix number of samples $m \geq Cs \log^4(s) \log(d)$
- ▶ Draw m samples $y_{\ell} = f(u_{\ell})$ according to Chebyshev measure
- ▶ Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be sampling matrix with entries $A_{\ell, \mathbf{k}} = T_{\mathbf{k}}(u_{\ell})$.
- ▶ Let $\mathbf{x}^{\#}$ be solution of

$$\min \|\mathbf{x}\|_{\omega, 1} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{m/s} \|f - f_{\Lambda_0}\|_{\omega, 1}$$

and set $f^{\#} = \sum_{\mathbf{k} \in \Lambda_0} x_{\mathbf{k}}^{\#} T_{\mathbf{k}}$.

Apply weighted ℓ_1 theory:

Consider a function

$$f = \sum_{\mathbf{k} \in \Lambda} x_{\mathbf{k}} T_{\mathbf{k}} \quad \text{on } [-1, 1]^d, \quad f \in \mathcal{A}_{\omega, p}$$

- ▶ Fix s , and fix $\Lambda_0 = \{\mathbf{k} \in \mathbb{N}^d : \omega_{\mathbf{k}}^2 \leq s\}$ **Reduced basis**
- ▶ Fix number of samples $m \geq Cs \log^4(s) \log(d)$
- ▶ Draw m samples $y_\ell = f(u_\ell)$ according to Chebyshev measure
- ▶ Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be sampling matrix with entries $A_{\ell, \mathbf{k}} = T_{\mathbf{k}}(u_\ell)$.
- ▶ Let $\mathbf{x}^\#$ be solution of

$$\min \|\mathbf{x}\|_{\omega, 1} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \sqrt{m/s} \|f - f_{\Lambda_0}\|_{\omega, 1}$$

$$\text{and set } f^\# = \sum_{\mathbf{k} \in \Lambda_0} x_{\mathbf{k}}^\# T_{\mathbf{k}}.$$

Then with probability exceeding $1 - e^{-d \log^3(s)}$,

$$\|f - f^\#\|_{L^\infty} \leq C_1 \left(\frac{\log^4(s) \log(d)}{m} \right)^{1/p-1} \|f\|_{\omega, p}$$

Limitations of weighted ℓ_1 approach

In the previous example,

$$N = |\Lambda_0| = e^d s^{2+\log_2(d)}$$

Weighted ℓ_1 minimization as a reconstruction method on such large scale problems is impractical

Fix $\tilde{s} \ll s$. Least squares projection onto

$$\tilde{N} = e^d \tilde{s}^{2+\log_2(d)}$$

is faster, but **too greedy**

Can we meet in the middle?

Summary

- ▶ We introduced weighted ℓ_1 minimization for stable and robust function interpolation, as taking into account both sparsity and smoothness present in natural functions of interest
- ▶ Along the way, we extended the notion of *restricted isometry property* to *weighted restricted isometry property*, a more mild condition that allows us to treat function systems with increasing $\|\cdot\|_\infty$ norm
- ▶ Weighted ℓ_1 minimization can overcome curse of dimension w.r.t. *number of samples* in high-dimensional approximation problems.

Extensions:

- ▶ We observe empirically that weighted ℓ_1 minimization is “faster” than unweighted ℓ_1 minimization. The steeper the weights, the faster. Justification?
- ▶ We observe that, with error on measurements $y = f(u) + \xi$, reconstruction results are similar, *provided* that the regularization parameter is chosen correctly in regularization methods. Equality constrained ℓ_1 minimization, weighted or not, leads to overfitting. Why?

References:

- ▶ Holger Rauhut and Rachel Ward. *Interpolation via weighted ℓ_1 minimization*. Arxiv, 2013.

Thank you!